Journal Paper

"Maintenance management based on Machine Learning and nonlinear features in wind turbines"

-Renewable Energy-

February 2020

Alfredo Arcos Jiménez
Ingenium Research Group, Castilla-La Mancha University, Spain
alfredo.arcosjimenez@uclm.es

Long Zhang
School of Electrical and Electronic Engineering, University of Manchester, UK
long.zhang@manchester.ac.uk

Carlos Quiterio Gómez Muñoz
Universidad Europea de Madrid, Spain
carlosquiterio.gomez@universidadeuropea.es

Fausto Pedro García Márquez
Ingenium Research Group, Castilla-La Mancha University, Spain
FaustoPedro.Garcia@uclm.es

Cite as: Jiménez, A. A., Zhang, L., Muñoz, C. Q. G., & Márquez, F. P. G. (2020). Maintenance management based on Machine Learning and nonlinear features in wind turbines. Renewable Energy, 146, 316-328.

DOI: https://doi.org/10.1016/j.renene.2019.06.135

Maintenance Management based on Machine Learning and Nonlinear Features in Wind Turbines

Alfredo Arcos Jiménez [1,a], Long Zhang [2,b], Carlos Quiterio Gómez Muñoz [3,c], Fausto Pedro García Márquez [1,d]

[1] Ingenium Research Group, Castilla-La Mancha University, Spain.
[a]alfredo.arcosjimenez@uclm.es, [d] faustopedro.garcia@uclm.es

[2] School of Electrical and Electronic Engineering, University of Manchester, UK
[b] long.zhang@manchester.ac.uk

[3] Universidad Europea de Madrid, Spain
[c]carlosquiterio.gomez@universidadeuropea.es

Abstract: Delamination is a common problem in wind turbine blades, creating stress concentration areas that can lead to the partial or complete rupture of the blade. This paper presents a novel delamination classification approach for reliability monitoring systems in wind turbine blades. It is based on the feature extraction of a nonlinear autoregressive with exogenous input system (NARX) and linear auto-regressive model (AR). A novelty in this paper is NARX as a Feature Extraction method for wind turbine blade delamination classification. Further, the NARX feature is demonstrated to be significantly better than linear AR feature for blade damage detection, and NARX can describe the inherent nonlinearity of blade delamination correctly. A real case study considers different levels of delamination employing ultrasonic guided waves that are sensitive to delamination. Firstly, the signals obtained are filtered and de-noised by wavelet transforms. Then, the features of the signal are extracted by NARX, and the number of features is selected considering the Neighbourhood Component Analysis as main novelties. Finally, six scenarios with different delamination sizes have been performed by supervised Machine Learning methods: Decision Trees, Discriminant Analysis, Quadratic Support Vector Machines, Nearest Neighbours and Ensemble Classification.

Index Terms: Wind Turbine Blade, Delamination, NARX, Machine Learning, Guided Waves.

NOMENCLATURE

AIC	Akaike Information Criterion
ANN	Artificial Neural Network
AR	Autoregressive
ARX	Autoregressive exogenous
AUC	Area Under Curve
DA	Discriminant Analysis
DT	Decision Tree
DTC	Decision Tree Complex
EBT	Ensemble Bagged Tree
FE	Features Extraction
FS	Features Selection
GW	Guided Wave
h-NLPCA	hierarchical Nonlinear Principal Component Analysis
KNN	K-Nearest Neighbours
MFC	Macro Fibre Composites
ML	Machine Learning
NARX	Nonlinear autoregressive with exogenous
NCA	Neighbourhood Component Analysis
NLPCA	Nonlinear Principal Components Analysis
PCA	Principal Components Analysis
QDA	Quadratic Discriminant Analysis
QSVM	Quadratic Support Vector Machine
ROC	Receiver Operating Characteristic
SHM	Structural Health Monitoring
SVM	Support Vector Machine
UGW	Ultrasonic Guided Waves
WKNN	Weighted KNN
WT	Wind Turbine
WTB	Wind Turbine Blade

1. INTRODUCTION.

A common problem in composite structures, such as wind turbine blades (WTB), is the appearance of delamination. The delamination is the disunion of one or more layers of the

composite material, which leads to a deterioration of the structural properties of the whole structure.

When a structure with any delamination is subjected to tensions and forces, e.g. in work regimen, stress concentration is generated at the edges of the disunion. It can generate the total breakage of the structure [1]. Due to the risks and large economic costs involved in the breakage of a WTB, a proper maintenance management based on predictive maintenance is required.

Structural health monitoring (SHM) is employed to analyse the condition of a structure. It is usually considered in a predictive and preventive maintenance strategy [2-4]. It is often used to detect failures in real time, or at fixed intervals by analysing the signals collected from sensors and using signal processing methods.

SHM systems have become an important research topic to reduce operation and maintenance cost. In the last decade, due to advances in the computational capacity of electronic systems and digital signal processing, SHM is having more relevance [5,6]. New SHM techniques and approaches have been developed for inspecting WTBs [7-9]. SHM is also employed to measure and evaluate the structural integrity in the new concept of "smart-blade", that involves the life cycle of the blade including design, operation, maintenance, repair and recycling [10]. Figure 1 shows the methodology of an "smart-blade".

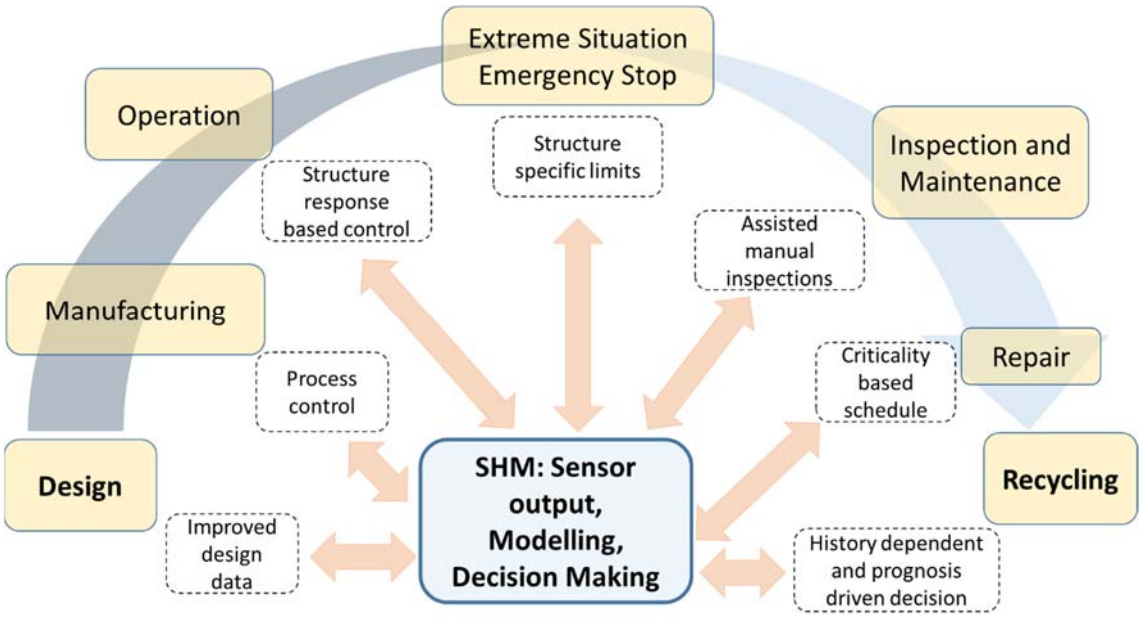


Fig. 1. Life stages of a wind turbine blade: "smart-blade" [10].

Techniques, such as Pattern Recognition from Machine Learning (ML), are applied to detect the different stages in the cycle of a WTB. Pattern recognition by ML has been used in many different issues. Pérez-Ortiz *et al*. perform a comprehensive review of problems and classification

algorithms with ML in the field of renewable energies [11]. Joshuva and Sugumaran present a compressive review on failure diagnosis methods and SHM for wind turbines [12]. N. Dervilis *et al*. applied pattern recognition to diagnose the failure of a WTB employing Principal Component Analysis (PCA), Non-Linear Principal Component Analysis (NLPCA) and Artificial Neural Networks (ANN) with Radial Basis Functions [13]. Panu Pratumnopharat *et al*. carried out a new approach using Meyer's Mexh, Daubachies, Morlet and Discrete Wavelets to obtain the historical stresses over the time in a wind turbine using vibration signals [14]. Altabey and Noori show a review of methodologies based on Lamb wavelets and ANN developed for the detection of damage in composite structures [15]. Feklistova and Hein identify the size and location of delamination in homogeneous and composite laminates using an aggregated approach that combines Haar wavelets, Support Vector Machine (SVM) and ANN [16]. Pavlopoulou *et al*. developed a work based on the SHM online of adhesive composite repairs, using ultrasonic guided waves (UGW) with post-processing techniques by Features Extraction (FE) and selection by PCA and NLPCA [17].

A general procedure in pattern recognition using ML is to filter the original signal, extract and select the relevant features, and classify the extracted features [18]. The classified results are used to identify WTB conditions or levels of damage severity [7,19,20]. SHM can reduce maintenance costs by preventing serious WTB failure, and can also increase profitability by reducing downtime [21,22].

There is extensive literature on the filtering and analysis of UGW. Standard statistical techniques have been employed for filtering [23]. Yu *et al.* [24] reduced noise using averaging techniques and Daubechies Wavelet Transform, filtering the high-frequency perturbations. Denoising and compression signal of UGW, based on the Discrete Wavelet Transform, was performed by Rizzo and di Scalea [25]. Hamming [26] presented a study about low-pass filters available for data smoothing.

FE techniques aims to create representative characteristics with lower dimensionality than the original data. FE can be extracted from models, such as Markov models and Autoregressive (AR) models [27,28]. The AR model is one of the most popular methods for FE of time-series dataset [29]. However, AR is a linear method, and a nonlinear relationship presented in the signal cannot be represented, and this may limit classification performance due to the lost nonlinear information.

This paper presents as novelty a Nonlinear Autoregressive with Exogenous input method (NARX) utilised as FE for detection and classification different delamination states in WTB. The features extracted by NARX presents better accuracy in the case study than linear AR features. The behaviour of Lamb waves propagation through defects is complex and nonlinear when they are

transmitted in anisotropic materials, e.g. WTB. NARX is a robust, simple and versatile algorithm for identifying non-linear dynamic systems, considering also Gaussian noises.

NARX was introduced by Leontaritis and Billings to describe a wide class on nonlinear systems [30]. NARX predicts the output value y_i by a non-linear function F from the input x_i, i.e. $y_i = F(y_{i-1} \cdots y_{i-n_y}; x_{i-1} \cdots x_{i-n_x})$. F is defined by regression approaches, e.g.: multinomial regression, non-parametric model forms based on automatic learning (Multi-layer Perceptron neural networks, MLP, Radio Basis Function, Gaussian Process, NARX models, etc.) [31].

NARX has been developed as a methodology in ML by applying a data-driven approach to prediction through the NARX Neural Network with exogenous inputs [32], e.g. predicting faults in gearboxes, bearings [33,34], predictive control [35] and forecast of wind speed and power [36]. The NARX model is often formulated as a linear combination of nonlinear functions to reduce computational costs referred as the linear-in-the-parameters model [37]. NARX models are properly structured for adaptive learning, and its structure setting is complex due to the number of terms grows rapidly.

The main selection methods are the Forward Regression Orthogonal Estimator, Randomized Model Structure Selection, Randomized Features Selection and Classifier, etc. PCA, Multidimensional Scaling, Kernel PCA, and Neighbourhood Component Analysis (NCA) are employed to reduce the dimensionality in classification learning [38]. NCA is proposed by Goldberger et al. for learning the Mahalanobis distance measurement for KNN classification algorithm [39]. Zhou et al. established a condition monitoring and fault diagnosis system to prevent abrupt bearing breaks during operation using NCA as a feature selection method [40].

Training process with supervised learning is employed to perform the classification [41]. Then, the trained classifier can be used for new data classification the classification performance is dependent on the choice of features and classifier models.

In this paper, five classifiers are employed to identify the delamination: Decision Trees, Discriminant Analysis, Support Vector Machines, Nearest Neighbours and K-Ensemble Classification. The results are compared with the AR-FE model that is obtained using the Yule-Walker method to verify the performance of the NARX features. Two conventional methods were used to evaluate the precision of the classifiers: Analysis Receiver Operating Characteristic (ROC) and Confusion Matrix. The classification of the different levels of delamination of the WTB has been performed employing: Wavelet Filtering, FE NARX, FS NCA and ML classifiers. Figure 2 shows the scheme of the proposed methodology.

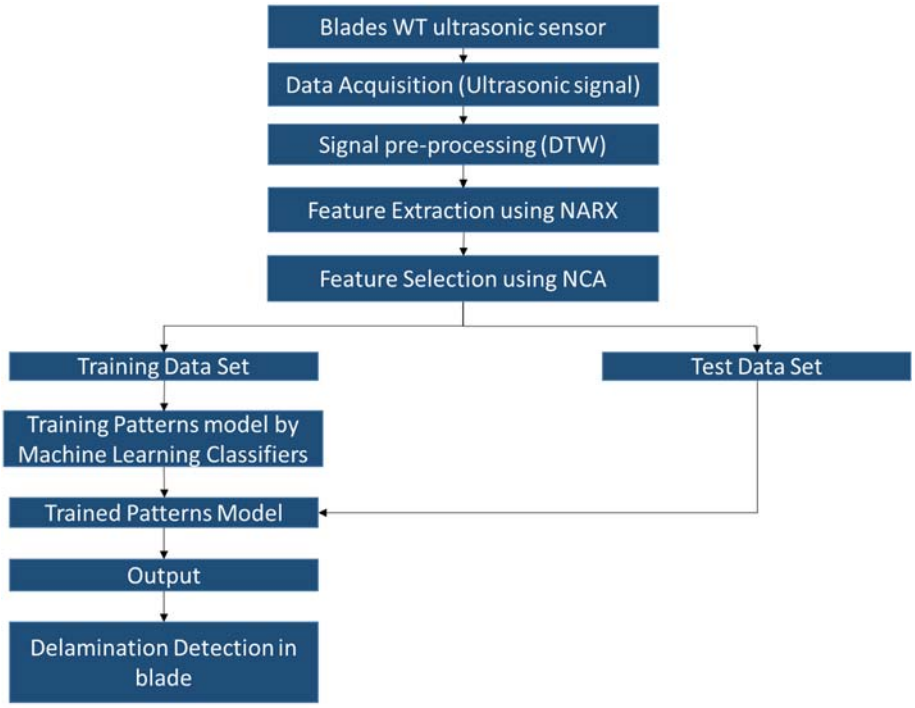


Fig. 2. Proposed Methodology.

Appendix B shows a summary of the state of the art showed in this paper.

2. CASE STUDY

2.1 Experimental procedure.

The experimental system is composed of a WTB section, an ultrasonic pulser-receiver system and ultrasonic transducers. This WTB is usually installed in wind turbines model G80-2.0 MW, being the Gamesa NACA 63.XXX + FFA-W3 model. It is made with fiber glass with a PVC foam core, and it is 39 meters length.

SHM of the WTB was done by UGW, transmitted through the composite material. The transmitter was excited with a sweep of frequencies to experimentally determine the best frequencies.

The transducer is Macro Fibre Composite (MFC) and it is composed of interdigitated piezoelectric fibers [42]. It can easily adapt to curved surfaces, e.g. the WTB surface. Figure 3 shows the sweep signals in the frequency domain, where the frequency range from 47 to 57 kHz provided better results. The ‘y’ axis is the power spectral density in decibels.

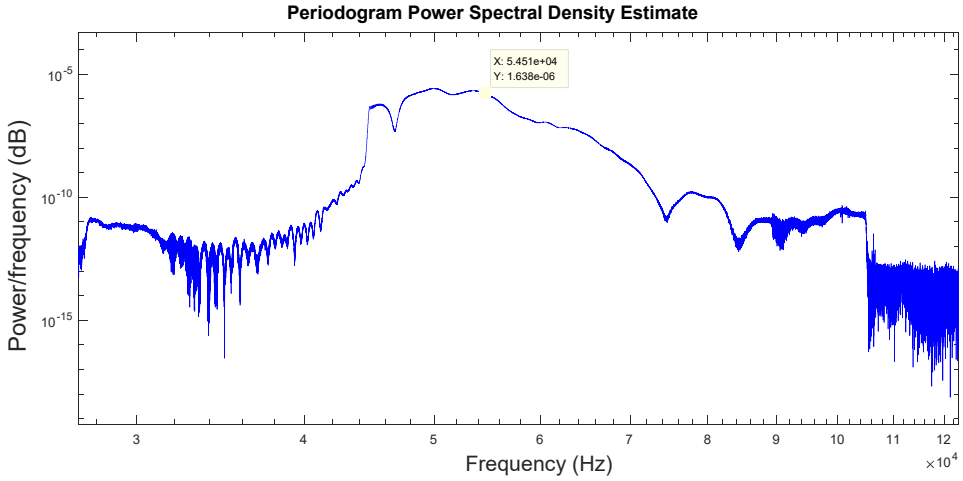


Fig. 3. Frequency sweep response for the wind turbine WTB and the MFC transducer.

The frequencies employed in this paper were: 25, 37, 55 and 100 kHz, being 37 and 55 kHz the frequencies that provided better results.

Six different case studies were carried out: the first case of study consists of transmitting UGW in the WTB without inducing any damage; the second is to induce internal damage of one centimetre in length and one centimetre in deep with a sharp instrument. The following cases are increased the depth of the defect in one centimetre up to 5 centimetres (Table 1). Table 1 shows all the scenarios considered in the experiments.

TABLE 1. DELAMINATION SIZE.

Scenario	'x' (cm)	Delamination Area(cm^2)
1	0 (Undamaged)	0
2	1	1
3	2	2
4	3	3
5	4	4
6	5	5

Figure 4 shows the transducer locations in the blade Gamesa NACA 63.XXX + FFA-W3 type, usually installed in wind turbines model G80-2.0 MW. The ultrasonic technique used was 'pitch and catch', where a short ultrasonic pulse is emitted by the MFC transmitter (Tx), and then it is collected by the MFC sensor (Rx). The ultrasonic pulse was a six-cycle Hanning pulse for each frequency.

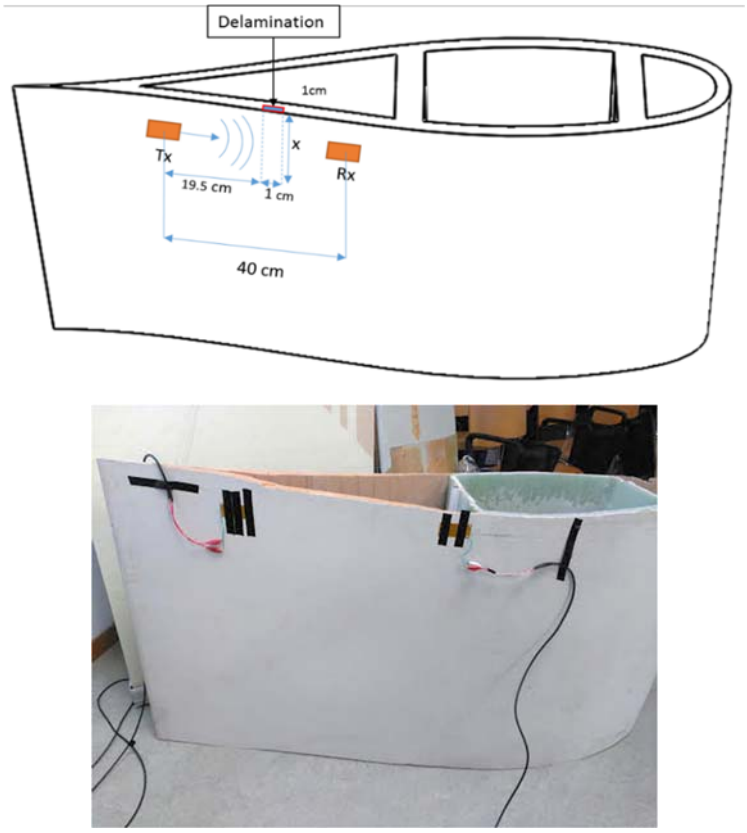


Fig. 4. Scheme of MFC location for delamination detection.

Six hundred experiments for each frequency were done. Figure 5 shows the signal for 55 kHz in each scenario with a sample rate of four million of samples per second.

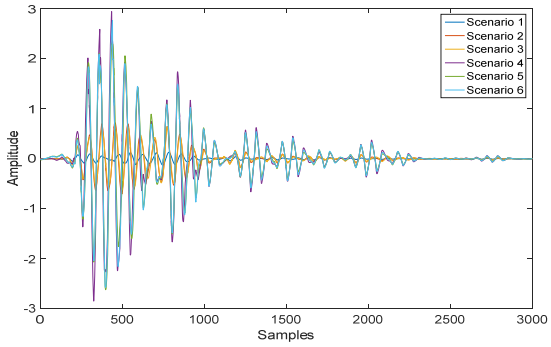


Fig. 5 Set of signals for different scenarios, collected for 55 kHz (Filtered).

2.2 Novel approach.

The signals are firstly filtered and de-noised by Wavelet transforms [43,44]. The signals contain noise that appears at high frequencies (random noise).

The Wavelet transform is a powerful method that allows to identify the local characteristics of a signal in the time and frequency domain. It presents some advantages that improves the limitations of resolution and the loss of information presented by the Short-Time Fourier Transform or the Fast Fourier Transform [45]. The Wavelet transform has several level decompositions. Each level decomposition is essentially a combination of two filters, where one is low pass filter, generating the filtered results that are called approximations, and other is high pass filter, producing the results called details [7].

In the case of the multi-level filters, they repeat the filtering process with the output signals from the previous level, leading the wavelet decomposition trees. Additional information is obtained by filtering at each level. However, more decompositions levels do not always mean better accurate results.

The Daubechies wavelet family were employed according to reference [46], where it is demonstrated that they are the most suitable for this type of signals because they are more sensitive to sudden changes, and they handle with boundary problems for finite length signals. The number of levels was set at seven after several experiments, where it was obtained the highest percentage of information. The lower wavelet approximation is removed from the original signal to avoid the trend and other undesirable components that appear in the low frequencies.

The denoising of the signals is performed employing a multilevel 1-D Wavelet analysis using the Daubechies family [47]. An overly aggressive filtering could eliminate information about the condition, e.g. small echoes that come from defects. The threshold for the de-noising is obtained by a wavelet coefficients selection rule using a penalization method provided by Birgé-Massart, which produces good results [48]. In contrast to other digital filters, the Wavelet de-noising filter does not produce an unwanted distortion of the signal characteristic parameters, e.g. time of flight.

Section 4 and 5 describes the FE of each ultrasonic signal extracted by AR and NARX. Section 5.B indicates the FS by AIC. The pattern recognition was done by ML through supervised learning. The results from the following classifiers were compared: Decision Trees, Linear Discriminant Analysis, Support Vector Machines, K-Nearest Neighbours and Ensemble Classification. The ROC and confusion matrix analysis has been employed to evaluate the classification and obtain the precision.

3. LINEAR FEATURES.

An AR system of extraction is employed to compare the linear and nonlinear regressive methodology. The current output is a linear combination of the past p outputs, considering also a white noise input, and it is set as an AR model of order p. The objective is to minimize the mean-square prediction error by setting the weights on the p past outputs of the auto-regression. If $y(t)$ is the current value of the output with a zero-mean white noise input, the AR(p) model can be written as equation (1),

$$y(t) = \sum_{i=1}^{p} \emptyset_i \, \mathrm{y(t-i)} + \xi(t) \tag{1}$$

where y(t) is the time series to be modelled, $\emptyset_i$ are the model coefficients, $\xi(t)$ is white noise that is independent from the previous points, and p is the order of the AR model.

Modified covariance method (forward-backward approach), correlation method (Yul-Walker approach), or covariance method, are widely used in AR [49]. In this paper, Yule-Walker has been applied to estimate the unknown model coefficients $\emptyset_i$ [50].

4. NONLINEAR FEATURES.

The NARX model is introduced in this section and then the details of the nonlinear FE procedure are discussed, followed by the feature number determination using NCA.

4.1 NARX model

A polynomial NARX model is used as the nonlinear function. The polynomial NARX model with an order up to 3 is given by equation (2) [51].

$$\begin{aligned} y(t) = \sum_{i_1=1}^{M} y(t-i_1)\,\theta_{i_1} + \sum_{i_1=1}^{M}\sum_{i_2=i_1}^{M} y(t-i_1)\,y(t-i_2)\theta_{i_1 i_2} \\ + \sum_{i_1=1}^{M} \cdots \sum_{i_l=i_{l-1}}^{M} y(t-i_2)\ldots y(t-i_2)\theta_{i_1 \ldots i_l} + \xi(t) \end{aligned} \tag{2}$$

where $y(t)$ are data sample at time *t,* t=1,...,N, and $y(t - i_1)$ are the delayed data with lagged $i_1, i_1 = 1, \dots, r$. The coefficient θ are the features. Nonlinear AR (NAR) is a special case of NARX where the input coefficients are zeros, i.e. the input has no contribution to the model.

Given a set of N training samples, equation (2) can be expressed in the matrix form as equation (3),

$$\mathbf{y} = \mathbf{P\Theta} + \mathbf{\Xi} \tag{3}$$

where $\mathbf{y} = [y(1), \dots, y(N)]^T$ is the coefficient vector, $\mathbf{\Theta} = [\theta_1, \dots \dots \theta_M]^T$ is the unknown parameter vector, $\mathbf{\Xi} = [\xi(1), \dots, \xi(N)]^T$ is the residual vector, and $\mathbf{P} = [\mathbf{p_1}, \dots, \mathbf{p_i}, \dots, \mathbf{p_M}]$ is a N-by-M matrix, being $\mathbf{p}_i = [\mathbf{p}_j(x(1), v_j), \dots, \mathbf{p}_j(x(N), v_j)]^T$ [52].

4.2 Features Extraction

The coefficients of NARX, given in equation (4), are obtained as:

Step 1. The matrix of first-order terms $(\sum_{i_1=1}^{M} x_{i_1}(t)\,\theta_{i_1})$ is

$$\mathbf{A}(M) = \begin{pmatrix} y_{(t)} & y_{(t-1)} & \cdots & y_{(t-M)} \\ \vdots & & & \vdots \\ y_{(N-1)} & y_{(N-2)} & \cdots & y_{(N-M)} \end{pmatrix} \tag{4}$$

when M = 3, then:

$$\mathbf{A}(3) = \begin{pmatrix} y_{(t)} & y_{(t-1)} & \cdots & y_{(t-2)} \\ \vdots & & & \vdots \\ y_{(N-1)} & y_{(N-2)} & \cdots & y_{(N-3)} \end{pmatrix} \tag{5}$$

Step 2. The matrix of second-order terms $(\sum_{i_1=1}^{M}\sum_{i_2=i_1}^{M} x_{i_1}(t)\,x_{i_2}(t)\theta_{i_1 i_2})$ is:

$$\mathbf{B}(M) = \begin{pmatrix} y_{(t)}y_{(t-1)} & \cdots & y_{(t)}y_{(t-M)} & & & \\ y_{(t+1)}y_{(t)} & \cdots & y_{(t+1)}y_{(t-M)} & y_{(t)}y_{(t)} & & \\ \vdots & & \vdots & & & \\ y_{(N-1)}y_{(N-2)} & \cdots & y_{(N-1)}y_{(N-M)} & y_{(N-2)}y_{(N-M)} & \cdots & y_{(N-M)}y_{(N-M)} \end{pmatrix} \tag{6}$$

For M=3,

$$\mathbf{B}(3) = \begin{pmatrix} y_{(t)}y_{(t-1)} & y_{(t)}y_{(t-2)} & & & & \\ y_{(t+1)}y_{(t)} & y_{(t+1)}y_{(t-1)} & y_{(t)}y_{(t)} & & & \\ y_{(N-1)}y_{(N-2)} & y_{(N-1)}y_{(N-3)} & y_{(t+1)}y_{(t+1)} & y_{(t)}y_{(t)} & y_{(t)}y_{(t)} \end{pmatrix} \quad (7)$$

There are two reasons for choosing the model with order M=3: The first reason is that the higher order terms have much less contributions than lower model terms. This can be illustrated using Taylor expansion, where the contributions of higher order model terms are negligible. Moreover, the higher order terms can easily approximate noises; the second reason is that the higher orders introduce significant model complexity. According to parsimonious principle, a simple model is always preferable to a complex one when they have similar model performance. Sometime, a model with higher order terms could be too complex and performs poorer than the simpler models. Therefore, majority practical application only used model order up to 3.

Step 3. For matrix of N variables, each term is the product of the N values $\sum_{i_1=1}^{M} \cdots \sum_{i_l=i_{l-1}}^{M} x_{i_1}(t) \ldots x_{i_l}(t)\theta_{i_1 \ldots i_l}$. For example, for a polynomial of 3 variables and M = 4, the terms are shown in Table 2.

TABLE 2. TERMS FOR THREE VARIABLES POLYNOMIAL AND M=4.

Nº	TERMS
1-5	$y_{(t-1)}$, $y_{(t-2)}$, $y_{(t-3)}$, $y_{(t-4)}$, $y^2_{(t-1)}$
6-8	$y_{(t-1)}$ $y_{(t-2)}$,$y_{(t-1)}y_{(t-3)}$,$y_{(t-1)}y_{(t-4)}$
9-12	$y^2_{(t-2)}$,$y_{(t-2)}y_{(t-3)}$,$y_{(t-2)}y_{(t-4)}$,$y^2_{(t-3)}$
13-16	$y_{(t-3)}y_{(t-4)}$,$y^2_{(t-4)}$,$y^3_{(t-1)}$,$y^2_{(t-1)}y_{(t-2)}$
17-19	$y^2_{(t-1)}$ $y_{(t-3)}$, $y^2_{(t-1)}y_{(t-4)}$,$y_{(t-1)}y^2_{(t-2)}$
20-21	$y_{(t-1)}y^2_{(t-3)}$,$y_{(t-1)}y^2_{(t-4)}$
22-23	$y_{(t-1)}y_{(t-2)}y_{(t-3)}$,$y_{(t-1)}y_{(t-2)}y_{(t-4)}$
24-25	$y_{(t-1)}y_{(t-3)}y_{(t-4)}$,$y^3_{(t-2)}$
26-28	$y^2_{(t-2)}y_{(t-3)}$,$y^2_{(t-2)}y_{(t-4)}$,$y_{(t-2)}$ $y^2_{(t-3)}$
29-31	$y_{(t-2)}y^2_{(t-4)}$,$y_{(t-2)}y_{(t-3)}y_{(t-4)}$,$y^3_{(t-3)}$
31-34	$y^2_{(t-3)}y_{(t-4)}$,$y_{(t-3)}y^2_{(t-4)}$,$y^3_{(t-4)}$

Step 4. **P** matrix consists of the matrices involved (**A**, **B**),

$$\mathbf{P} = ((\mathbf{A}) \quad (\mathbf{B})) \quad (8)$$

Step 5. The solution of equation (3) is given by equation (8).

$$\boldsymbol{\Theta} = \mathbf{P}^{-1}\mathbf{y} \tag{9}$$

where $\boldsymbol{\Theta} = [\theta_1, \ldots . \theta_M]^{\mathrm{T}}$ is the extraction features parameter vector. The direct inverse of $\mathbf{P}$ matrix can be used to solve the model parameters. Alternatively, least squares methods are often used to compute the model coefficients by using the equation $\boldsymbol{\Theta} = (\mathbf{P}^{\mathrm{T}}\mathbf{P})^{-1}\mathbf{P}^{\mathrm{T}}\mathbf{Y}$. The least squares method tends to be more popular as it is more numerically robust than direct inverse operation.

4.3 Feature Selection

NCA has been applied to reduce the dimensionality in AR and NARX [38,40]. Let $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_{2,\ldots,}\mathbf{y}_{\mathrm{n}}\}$ be labelled features set, where $\mathbf{y}_{\mathrm{i}} \in \mathbb{R}^{\mathrm{d}}$, where the corresponding class labels are $\{\mathbf{c}_1, \mathbf{c}_{2,\ldots,}\mathbf{c}_{\mathrm{n}}\}$. The objective is to find a vector of weighting *w* to select the subset of characteristics optimizing the classification of nearest neighbours. *w* is the weighted distance between two samples $\mathbf{y_i}$ and $\mathbf{y_j}$, given by equation (10):

$$D_{\mathrm{w}}(\mathbf{y_i}, \mathbf{y_j}) = \sum_{l=1}^{d} w_l^2 |y_{il} - y_{jl}| \tag{10}$$

where w_l is a weight associated with the l^{th} feature.

The reference point in NCA is chosen randomly, where all points in $\mathbf{X}$ have a probability of being selected as the reference point. The point x_i selects the point x_j as its neighbour with a probability p_{ij}, receiving the class label from the neighbour selected. p_{ij} is defined by equation (11):

$$\begin{cases} p_{ij} = \dfrac{\mathcal{K}\left(D_{\mathrm{w}}(y_i, y_j)\right)}{\sum_{k \neq i} \mathcal{K}\left(D_{\mathrm{w}}(y_i, y_k)\right)}; if\ i \neq j \\ \\ p_{ij} = 0; if\ i = j \end{cases} \tag{11}$$

where $\mathcal{K}(z) = \exp(-\frac{z}{\sigma})$ is a Kernel function, and the Kernel with σ is an input parameter that influences the probability of each point to be selected as the reference point. The probability of the query point x_i to be correctly classified, p_i, is given by equation (12):

$$p_i = \sum\nolimits_j C_{ij}\, p_{ij} \tag{12}$$

307 being $C_{ij} = 1$ if the class label $C_i = C_j$, and $C_{ij} = 0$ if the class label $C_i \neq C_j$

308 The true leave-one-out classification accuracy $\xi(w)$ is calculated by equation (13):

$$\xi(w) = \sum_i \sum_j C_{ij}\, p_{ij} - \lambda \sum_{l=1}^{d} w_l^2 \tag{13}$$

309 where λ is a regularization parameter set by cross validation. The regularization term $\lambda \sum_{l=1}^{d} w_l^2$ is
310 introduced in equation (12) to improve the feature selection and avoid overfitting.

311 A gradient rule is applied by differentiating $\xi(w)$ regarding to w_l in equation (14):

$$\frac{\partial \xi(w)}{\partial w_l} = 2\left(\frac{1}{\sigma}\sum_i \left(p_i \sum\nolimits_{j \neq i} p_{ij} |y_{il} - y_{jl}| - \sum\nolimits_j C_{ij}\, p_{ij} |y_{il} - y_{jl}| - \lambda\right) w_l\right) \tag{14}$$

312

313 **5. CLASSIFICATION METHODS**

314 ML multiclass with supervised learning is employed to classify the scenarios of the WTB
315 conditions. The classifiers used are: Decision Trees (DT); Quadratic Discriminant Analysis
316 (QDA); Support Vector Machines (SVM); K-Nearest Neighbours (KNN), and; Ensemble
317 Classification.

318 DT is a classifier used because of a complex decision process is segregated in simpler decision
319 processes [53]. DT consists in three stages: Construction of the maximum tree using a binary
320 partition procedure; tree pruning, and; selection of the optimal tree by means of a cross-validation
321 procedure. A classification tree is graphically represented by nodes and branches. The tree is
322 initially represented by the root node.

323 Construction of the maximum tree is based on the cyclic decomposition of an initial data group,
324 called the parent group, into two mutually exclusive subgroups, namely daughter groups. The
325 subgroups are configured to increase the homogeneity of the new group using the impurity
326 function.

327 The impurity function determines the quality of a node. The impurity, or partition criterion, is
328 given by Gini's Diversity Index (GDI), equation (15):

$$GDI = 1 - \sum_i p^2(i) \tag{15}$$

329 being $p(i)$ the observed fraction of classes with class i that has the node. A node with only one
330 class, named pure node, $GDI = 0$, being $GDI > 0$ in other cases.

The optimal tree generally presents overfitting [54]. The second phase of the process is to reduce the size of the tree, that consists of cutting off terminal nodes until the optimal size. The tree is divided in different subtrees, that are compared to find the optimal, considering the measure of cost complexity $R_\alpha(T)$, given by equation (16):

$$R_\alpha(T) = R(T) + \alpha\left|\tilde{T}\right| \tag{16}$$

being $R(T)$ the average within-node sum of squares, $\left|\tilde{T}\right|$ is the tree complexity, i.e. the total number of nodes of the subtree, and α the complexity parameter employed as a penalty for each additional terminal node. α is gradually increased from 0 to 1 in the pruning procedure, and it is considered the value that minimize $R_\alpha(T)$. The optimal subtrees are selected according to the minimal predictive error models employing cross validation. The advantage of this classifier is the flexibility and ability to be applied to nonlinear relationships between features and classes.

QDA is a common technique for multivariate analysis, based on finding the quadratic combination of variables, optimising the classification of differentiate classes by means of a discriminant function [55].

The classification of a discriminant analysis process can be summarized as:

a) To calculate the prior probability π_i of the class *i*, i.e. the expected proportion of observations belong to each group.

b) To determine if the variance or covariance matrix is homogeneous in all groups. It is assumed that the vector of feature variables **X** is multivariate normally distributed in the group with mean $\boldsymbol{\mu_i}$ and covariance matrix $\sum_{\boldsymbol{i}}$.

c) To estimate the necessary parameters for the conditional probability functions by equation (17),

$$f_i(\mathbf{X}) = \frac{1}{(2\pi)^{P/2}|\sum_i|^{1/2}}\, exp\left[-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu}_i)^T \sum_i^{-1}(\mathbf{X}-\boldsymbol{\mu}_i)\right] \tag{17}$$

where *P* is a dimension factor (2 for QDA), and *T* is a transpose operator.

d) To calculate the result of the discriminant function by equation (18) to set the group assigned

$$g_i(X) = -\frac{1}{2}(X-\mu_i)^T \sum_i^{-1}(X-\mu_i) - \frac{1}{2}\, log(|\sum_i|) + log(\pi_i) \tag{18}$$

e) To use cross-validation to estimate the probabilities of misclassifications.

SVM considers each feature as a point in a high-dimensional space, where the number of dimensions is equal to the number of rating levels [56]. An optimal hyperplane is defined to separate the dataset based on class membership. The Kernels function $K(x_i, x') = (\phi(x_i), \phi(x'))$ is used to transform the input data to a higher dimensional space, where a decision boundary can be designed. Given n training vectors, $\boldsymbol{x_i} \in \Re^p : i = 1, \ldots n$, defined by a set of p descriptive variables $\boldsymbol{x_i} = \{x_{i1,\ldots,} x_{ip}\}$, by the class label $y_i \in \{-1, +1\}$, and using the Lagrange function, the optimization problem is solved by the decision function (19),

$$D(x) = w\phi(\mathbf{x_i}) + b \tag{19}$$

where w and b are the SVM parameters, and $\phi(x)$ is a kernel function, and the quadratic function Kernel is $K(\mathbf{x_i}, \mathbf{x}') = \left((\mathbf{x_i}, \mathbf{x}') + 1\right)^2$.

The hyperplane is defined by equation (20), and the distance between the hyperplane and pattern **x** is given by equation (19).

$$\frac{D(\mathbf{x})}{\|\mathbf{w}\|} \tag{20}$$

A training classifier is used to find the value w that maximizes the margin between the class boundary and the training patterns [56]. The objective function, J, of the training algorithm is given by equation (21).

$$J = \frac{\|\mathbf{w}\|^2}{2} \tag{21}$$

One vs one is a method used in pattern recognition when multiple patterns exist [57].

KNN classifier has been used for pattern classification and ML [58]. KNN is based on the determination of the nearest neighbours and to set the class using the neighbours. Given a set training observations **X** of n feature measurements, $\mathbf{X} = (x_1, \cdots, x_N)^T$, and the class label **Y** = 1,…,C, a KNN classifier allows to find the K nearest neighbours of a query point x_0 in **X,** and then it predicts the class label of x_0. Given a query x_0, its unknown class y_0 is assigned by two steps: Firstly, a set of k similar labelled target neighbours for the query x_0 is identified. The Euclidian distance is given by equation (22) for a set $T' = \{x_i^{NN}, y_i^{NN}\}_{i=1}^{k}$, set in a growing order in terms of Euclidean distance $d\left(x_0, x_i^{NN}\right)$ between x_0 and x_i^{NN}.

$$d\left(\mathrm{x}_0, x_i^{NN}\right) = \sqrt{\left(\mathrm{x}_0 - x_i^{NN}\right)^T \left(\mathrm{x}_0 - x_i^{NN}\right)} \tag{22}$$

383

384 Secondly, the class label of the query point is predicted by the majority of voting of its nearest
385 neighbours, given by equation (23)

$$y_0 = \arg\max_y \sum_{(x_i^{NN}, y_i^{NN}) \in T'} \delta\left(y = y_i^{NN}\right) \tag{23}$$

386 where y is the class label, y_i^{NN} is the class label for i-th nearest neighbour among its k nearest
387 neighbours. The Dirac delta function is expressed by equation (24),

$$\delta\left(y = y_i^{NN}\right) = \begin{cases} 1 \; if \; y = y_i^{NN} \\ 0 \; otherwise \end{cases} \tag{24}$$

388

389 Weighted KNN (WKNN) classifier has been used in this paper, where the nearest neighbours
390 outweigh the farthest. The nearest neighbours have a higher weight than the farther ones in
391 WKNN. Consequently, the weighted distance function shall be used, where the ω_0^i weight for the
392 i-th nearest neighbour of the query x_0 is defined by equation (25).

$$\omega_0^i = \begin{cases} \frac{d\left(\mathrm{x}_0, x_k^{NN}\right) - d\left(\mathrm{x}_0, x_i^{NN}\right)}{d\left(\mathrm{x}_0, x_k^{NN}\right) - d\left(\mathrm{x}_0, x_1^{NN}\right)}, if \;\; d\left(\mathrm{x}_0, x_k^{NN}\right) \neq d\left(\mathrm{x}_0, x_1^{NN}\right) \\ 1 \qquad\qquad\qquad\qquad\quad , if \;\; d\left(\mathrm{x}_0, x_k^{NN}\right) = d\left(\mathrm{x}_0, x_1^{NN}\right) \end{cases} \tag{25}$$

393 where the result of the classification of the vote, based on the weighted majority vote, is obtained
394 by equation (26)

$$y_0 = \arg\max_y \sum_{(x_i^{NN}, y_i^{NN}) \in T'} \omega_0^i \times \delta\left(y = y_i^{NN}\right) \tag{26}$$

395 Equation (26) indicates that a neighbour with less distance has a weight greater than one with
396 more distance. Therefore, the nearest neighbour will have a weight of 1, and the furthest
397 neighbour will have a weight of 0. The neighbours between these distances will have linear scaled
398 weights.

399 Ensemble classification methods are learning algorithms that construct a set of classifiers whose
400 individual decisions are combined generally by weighted, or unweighted voting, to classify the
401 set of features of each pattern [59].

402 Bagging, boosting, and random subspaces are general techniques that can be used with any type
403 of base classifier. In this case, Ensemble Bagged Tree (EBT) was the ensemble with the best

results. EBT depends of the number of splits and the number of learners. The complexity of the tree increases with the number of splits. An overfitting is done when the number of splits is not correct. The accuracy is proportional to the number of learners, but it can be time consuming to fit. A high predictive power can be got by a few hundreds of learners. It is necessary to carry out different tests to set the hits percentage for each algorithm. EBT uses the Breiman's "random forest" algorithm [60], i.e. a combination of tree predictors, where each tree depends of a random vector sampled independently. The distribution of the trees in the forest is the same. The size of the trees in the forest is limited by the generalization error for forests convergence. This error depends on the correlation between the trees and their strength.

5.1 Evaluation

ROC, as the confusion matrix, has been employed to evaluate the classifications, obtaining the precision of classifier, or tape rate, *P,* by equation (27),

$$P = tp/(tp + fn) \tag{27}$$

where tp are true positives correctly classified, and fn are false negatives in the confusion matrix.

Macro- averaging and micro-averaging are used to evaluate the performance average.

Macro-average (P_i^{Ma}): P^{Ma} is obtained by the average over all P_i^{Ma}, where *Ma* denotes Macro-average, and *i* the scenario. They are then calculated for each category, i.e. the values precision is evaluated locally P_i^{Ma}, and then globally P^{Ma}.

Micro-average (P_i^{μ}): The precision value is obtained as: i) tp_i, fp_i, fn_i are calculated for each scenario, where *fp* are false positives in the confusion matrix; ii) the value of *tp*, *fp* and *fn* are calculated as the sum of tp_i, fp_i, fn_i, and; iii) to apply the definition of the measure that corresponds to it.

There are several indices extracted from ROC to evaluate the classifier efficiency. Area Under Curve (AUC) [61] is the area between the ROC curve and the negative diagonal [62]. AUC is optimized by equations (28) and (29).

$$AUC = (P - fprate + 1)/2 \tag{28}$$

$$fprate = fp/(fp + fn) \tag{29}$$

5.2 Comparative analysis of classifiers

The Friedman test, recommended by Demšar [63], or its modification made by Garcia and Herrera [64], are employed to perform the comparative analysis of different classifiers. The Friedman test

assumes that all classifiers can achieve the same performance with the same classification accuracy. It is called the null hypothesis [65]. The Bonferroni-Dunntest [66] is used to determine any significant differences between the top-ranked classifiers, and, finally, the Holm Test [67] is applied to contrast the results. The T-test [68] has been applied to confirm the null hypothesis and validate the results.

6. RESULTS

6.1 Selection Features

The signal has been analysed where it provides more information about the fault, i.e. in the second third. The feature parameter vector Θ has been obtained to reduce the number of features, i.e. to avoid "the curse of the dimensionality". 15 features have been used for AR and Yule Walker, and 30 for NARX (the first 15 features are of the first order). The number of features has been selected by NCA, and the results for NARX are shown in Figure 7.

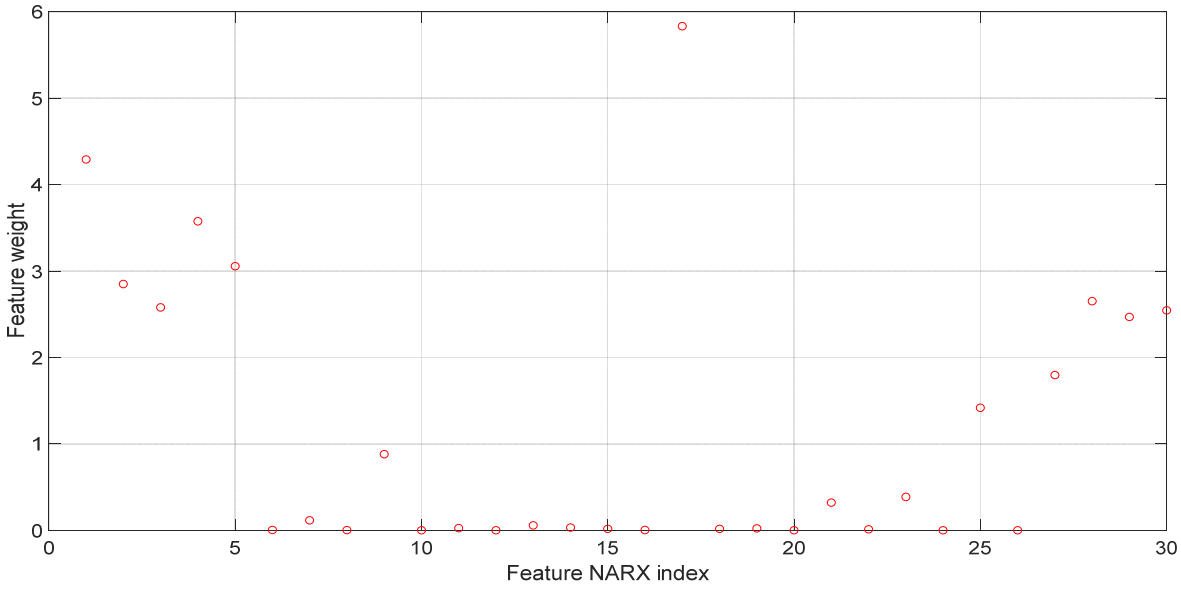


Fig. 6 NCA Features.

According to the results given by NCA, the value p = 15 has been selected.

6.2 Analysis of the results in terms of measurement precision

The classification is done when the FE is obtained by NARX and AR methods and the features number is selected. The FS vector of each signal has been introduced in different classifiers. The FS number p =15 is the number of inputs in the classifier. The NARX and AR methods are compared with different classifiers, see section VI, and the results are shown in Table 3. The first

main column describes the scenarios and the precision parameters; the remaining columns show different classifiers, and each column is subdivided into three sub-columns. The first sub-column shows the FE results employing AR by Yule-Walker. The second corresponds to the NARX results. The last sub-column indicates an improvement in the NARX results compared to the results given by AR. The values of micro and macro precision accuracies are given in the seven and eighth rows. The coefficient given by the Friedman tests and their classification are shown in the last rows.

The FE provided by NARX gives better results than AR in all scenarios, see Table 3.

The Demšar method has been employed to identify the significant difference between the classifiers and to establish the ranking between them.

The test results for significant differences between classifiers and for each experiment show that the Friedman test for AR did not reject the null hypothesis (p-value = $7{,}8894^{-04} \leq 0{,}05$). The Bonferroni-Dunn test, see Figure 8, rejects the null hypothesis for $p \leq 0{,}05$, with a confidence value α=0,05 for the classifier 1 (DTC). The Holm test does not reject the null hypothesis for the classifier.

The NARX Friedman test did not reject the null hypothesis because no significant differences were found (*p-value*=0.0042 ≤ 0.05). The Bonferroni-Dunn test rejects the null hypothesis with $p \leq 0{,}05$, with a confidence value α=0,05 for classifier 1 (DTC), while the test Holm does not reject the null hypothesis.

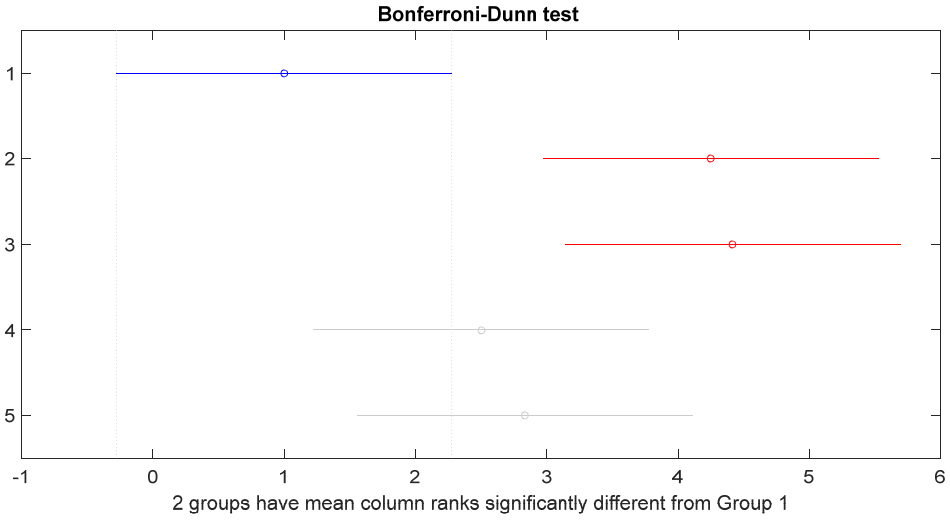


Fig. 7 Bonferroni-Dunn test for AR and NARX with α=0,05.

The t-Test is applied for NARX (Table 3) to validate the results between the classifiers. The test rejects the null hypothesis in the cases of the DT-QDA, DT-QSVM and DT-EBT classifiers. Consequently, DT can be discarded for this purpose. For QDA-QSVM, QDA-WKNN, QDA-

EBT, QSVM-WKNN, QSVM-EBT, WKNN-EBT, the p-value is ≥ 0.05, i.evalidating these methods.

TABLE 3 T-TEST: NARX.

	DT	QDA	QSVM	WKNN	EBT
DT		0.02911	0.03264	**0.77067**	0.01856
QDA	0.02911		**0.36089**	**0.09865**	**0.07600**
QSVM	0.03264	**0.36089**		**0.11716**	**0.09254**
WKNN	**0.77067**	**0.09865**	**0.11716**		**0.17900**
EBT	0.01856	**0.07600**	**0.09254**	**0.17900**	

Table 4 shows the T-Test for AR. It does not reject the null hypothesis for QDA-QSVM, QDA-WKNN, QDA-EBT and EBT-WKN classifiers. The t-Test indicates concordance of results between QSVM-QDA regarding to the classifiers. DT is also discarded in AR due to its low concordance with the classifiers.

TABLE 4 T-TEST: AR.

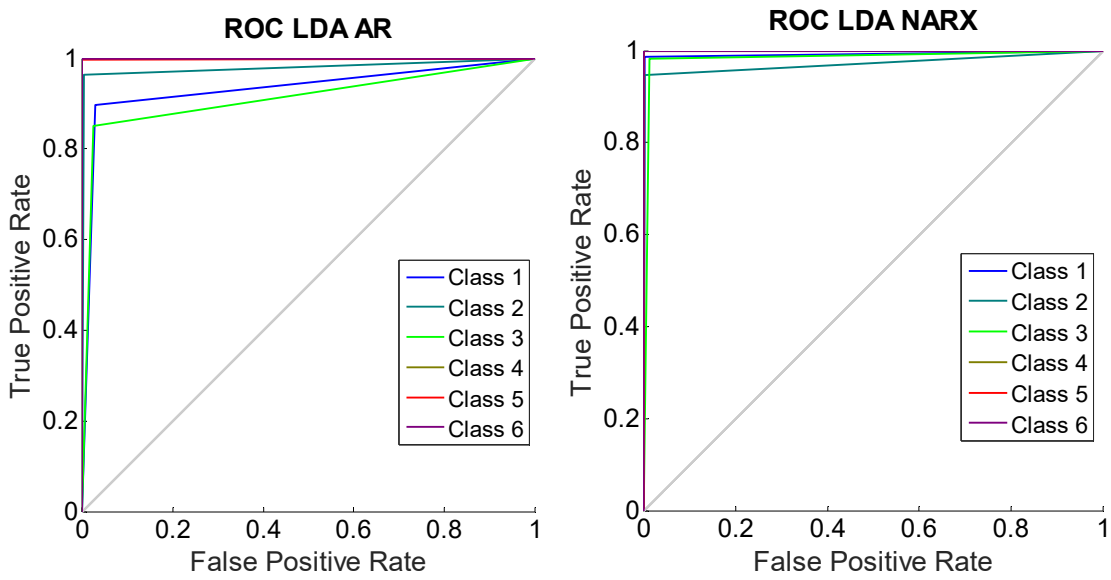
	DT	QDA	QSVM	WKNN	EBT
DT		0.00442	0.00478	0.00027	0.00044
QDA	0.00442		**0.41848**	**0.10283**	**0.17127**
QSVM	0.00478	**0.41848**		0.02674	0.03711
WKNN	0.00027	**0.10283**	0.02674		**0.16222**
EBT	0.00044	**0.17127**	0.03711	**0.16222**	

QSVM is the best classifier for AR (Table 5). Finally, the t-Test indicates that the QDA classifier is the best classifier for NARX. The classifiers for NARX FE do not present significant difference in terms of precision according to the evaluation tests, i.e. they present accuracy and robustness.

6.2 Analysis of the results employing AUC.

Table 4 shows the AUC results, where all scenarios exceed 0.87 (87%) of successes in the AR FE, and 0.92 (92%) in the NARX FE. Furthermore, it is observed that QSVM and LDA are close to 1 (100%) in all scenarios.

Figure 9 shows the results of the best classifier with AR FE (a) and NARX FE (b). It shows that, for LDA NARX, all the curves are closer to one compared to the curves for LDA AR.



(a) (b)

Fig. 8 Curve ROC Classification: LDA AR (a) and LDA NARX (b)

The test results for AUC are close to the previous sections, i.e. the Friedman test, as in post-hoc tests, shows that the QDA classifier is the best for the E-1; the Holm test confirms that the QDA classifier gives better results.

7. CONCLUSIONS

The paper shows a novel approach to determine delamination in a wind turbine blade employing ultrasonic guided waves and classifiers. A Nonlinear Autoregressive with Exogenous input system (NARX) method has been introduced for Feature Extraction.

The FE NARX method is set by Receiver Operating Characteristic curves, with the measure of the Area Under Curve and the Precision Measure values, where a set of features is selected considering the Neighbourhood Component Analysis method. Each scenario has been performed by Machine Learning through supervised learning. The classifiers used were: Decision Trees, Quadratic Discriminant Analysis, Quadratic Support Vector Machines, K-Nearest Neighbours and Ensemble Classification. Macro-averaging and micro-averaging are employed to evaluate the performance average across categories. FE by NARX improves the performance of all classifiers with respect to Autoregressive, according to the Precision Measure and Area Under Curve parameters.

The approach is able to differentiate between 6 different levels of delamination. The ultrasonic signals were experimentally collected in an actual wind turbine section. Different classifiers were used to identify the delamination size. The recommendations for comparative analysis of

classifiers by Demšar, and the extensions by Garcia and Herrera, have been employed. The Friedman test was used to confirm the null hypothesis. The Bonferroni-Dunn test was applied to determine any significant differences between the top-ranked classifier and, finally, the Holm Test was employed to contrast the results. The NARX method improves the AR between 3.5% and 7.6%.

Acknowledgement

The work reported herewith has been financed by the Spanish Ministerio de Economía y Competitividad, under Research Grant Ref.: DPI2015-67264-P.

Appendix A. Results Precision

TABLE 5. RESULTS PRECISION

SCENARIO	DECISION TREE COMPLEX			QDA			QSVM			WEIGHTED KNN			ENSEMBLE BAGGED TREE		
	P_i (AR)	P_i (NARX)	IMPROVE (%)	P_i (AR)	P_i (NARX)	IMPROVE (%)	P_i (AR)	P_i (NARX)	IMPROVE (%)	P_i (AR)	P_i (NARX)	IMPROVE (%)	P_i (AR)	P_i (NARX)	IMPROVE (%)
1	0.827	0.915	8.8	0.857	0.995	13.8	0.913	0.968	5.5	0.872	0.865	-0.7	0.898	0.955	5.7
2	0.813	0.900	8.7	0.982	0.992	1.0	0.963	0.970	0.7	0.855	0.947	9.2	0.878	0.945	6.7
3	0.767	0.867	10.0	0.873	0.953	6.0	0.943	0.965	2.2	0.835	0.857	2.2	0.840	0.927	8.7
4	0.92	0.987	6.7	0.997	1.000	0.3	0.992	1.000	0.8	0.970	1.000	3.0	0.958	1.000	4.2
5	0.963	0.97	0.7	1.000	1.000	0.0	1.000	1.000	0.0	0.998	0.995	-0.3	0.997	0.997	0.0
6	0.893	1.000	10.7	1.000	1.000	0.0	0.982	1.000	1.8	0.928	1.000	7.2	0.948	1.000	5.2
P^{μ}	0.864	0.940	7.6	0.951	**0.990**	3.9	**0.966**	0.984	1.8	0.909	0.944	3.5	0.920	0.971	5.1
P^{Ma}	0.864	0.940	7.6	0.951	**0.990**	3.9	**0.966**	0.984	1.8	0.909	0.944	3.5	0.920	0.971	5.1
RANKING	1	1.67		4.25	4.17		4.41	4.00		2.50	2.25		2.83	2.92	
CLASSIFIC.	**5**	**5**		**2**	**1**		**1**	**2**		**4**	**4**		**3**	**3**	

TABLE 6. RESULTS BY AUC

SCENARIO	DECISION TREE COMPLEX			QDA			QSVM			WEIGHTED KNN			ENSEMBLE BAGGED TREE		
	P_i (AR)	P_i (NARX)	IMPROVE (%)	P_i (AR)	P_i (NARX)	IMPROVE (%)	P_i (AR)	P_i (NARX)	IMPROVE (%)	P_i (AR)	P_i (NARX)	IMPROVE (%)	P_i (AR)	P_i (NARX)	IMPROVE (%)
1	0.917	0.945	2.8	0.940	0.993	5.30	0.969	0.989	2.0	0.948	0.946	-0.20	0.940	0.977	3.70
2	0.876	0.937	6.10	0.981	0.973	-0.80	0.974	0.989	1.5	0.907	0.949	4.20	0.912	0.973	6.10
3	0.903	0.919	1.60	0.918	0.985	6.70	0.954	0.973	1.9	0.902	0.926	2.40	0.931	0.955	2.40
4	0.952	0.991	3.90	0.999	1.000	0.10	0.993	1.000	0.7	0.968	0.993	2.50	0.972	1.000	2.80
5	0.989	0.980	-0.90	0.998	0.999	0.10	0.996	0.993	-0.3	0.989	0.990	0.10	0.996	0.997	0.10
6	0.951	1.000	4.90	1.000	1.000	0.00	0.993	1.000	0.7	0.976	1.000	2.40	0.972	1.000	2.80
RANKING	1.25	1.33		4.25	4.16		4.25	3.83		2.25	2.16		3.00	3.50	
CLASSIFIC.	**5**	**5**		**1**	**1**		**1**	**2**		**4**	**4**		**3**	**3**	

Appendix B. Signal, Feature and classification processing

Paper	Task	Method	Description
[14]	Signal Processing	Meyer's Mexh, Daubachies, Morlet and Discrete Wavelets	Historical stresses over the time in a wind turbine using vibration signals
[15]	Signal Processing	Lamb wavelets	Review of methodologies developed for the detection of damage in composite structures
[16]	Signal Processing	Haar wavelets	Identification of size and location of delamination in homogeneous and composite laminates
[23,69]	Signal Processing. Filtering	Daubechies Wavelet Transform	Review of standard statistical techniques employed for filtering
[26]	Signal Processing. Filtering	Low-pass filters	Digital Filters
[27]	Signal Processing. Filtering	PCA	Approach for detecting and identifying faults in railway infrastructure components
[24]	Signal Processing. Denoising	Wavelet denoising and Hilbert Transform	Development of an integrated ultrasonic structural radar algorithm for the use of active piezoelectric wafers to detect defects over a large area of a thin plate sample.
[25]	Signal Processing. Denoising	Discrete Wavelet Transform	Monitoring of multi-wire steel strands by the use of ultrasonic guided waves, through the implementation of the Discrete Wave Transform.
[13]	Feature Extraction and Selection	Principal Component Analysis (PCA), Non-Linear Principal Component Analysis (NLPCA)	Pattern recognition to diagnose the failure of a WTB
[17]	Feature Extraction and Selection	Principal Component Analysis (PCA), Non-Linear Principal Component Analysis (NLPCA)	Developed a work based on the SHM online of adhesive composite repairs
[18]	Feature Extraction	Autoregressive (AR) Yule-Walker equation	Detection and localization the delamination WTB
[21]	Feature Extraction	Autoregressive (AR) Yule-Walker equation	Determine effectively the temperature of pipe in Concentrate Solar Plant
[70]	Feature Extraction and Selection	Linear and Non-Linear Method	Review of methods developed in the field of damage detection, diagnosis, forecasting and control in flight vehicles.
[28]	Feature Extraction	PCA	A method of pattern recognition based on grouping by linear relation a set of railway infrastructure faults.
[71]	Feature Extraction	AR model	Review on machinery diagnostics and prognostics implementing condition-based maintenance
[72]	Feature Extraction	AR	Structural health monitoring techniques of pattern recognition of two different structural conditions of a boat.
[29]	Feature Extraction	AR and Variational mode decomposition	Seizure detection techniques for epileptic electroencephalogram
[30]	Feature Extraction	ARM, ARX and NARMAX models	Recursive input-output models for non-linear multivariate discrete-time systems
[52]	Feature Extraction	NARMAX	Input-output representations of non-linear discrete-time systems
[31]	Feature Extraction	NARMAX, NARX, Gaussian Process	GP-NARX framework by providing a means of converting the GP prediction bounds in the time domain into bounds on the hfrfs
[32]	Feature Extraction	NARX-NN, wavelet, time delay neural network,	Determination of remaining useful life (RUL) of bearings
[33]	Feature Extraction and Forecasting	Adaptive neuro-fuzzy inference system (ANFIS) and NARX	Predicting gearbox health condition index extracted from the vibration signals
[34]	Feature Extraction and Forecasting	Neuro-fuzzy inference system, Nonlinear Autoregressive model,ANN	Prognostic techniques applied to maintenance of wind turbines(Review)
[35]	Feature Extraction and Forecasting	Multi-objective optimization, Neural network prediction	Wind power ramping events and reduce WT load on the blades
[36]	Forecasting	Neural Networks	Multi-step ahead wind speed and power forecasting
[37]	Feature Extraction and Classification	NARX, Decision tree, SVM, K-Nearest Neighbours	Pattern recognition through guided waves and Machine Learning for ice detection
[73]	Feature Extraction	NARX, Support vector regression	Overview of basic research on model selection approaches for linear-in-the-parameter models
[74]	Feature Extraction	Orthogonal parameter estimation algorithm	Unbiased estimation in the presence of correlated noise
[75]	Feature Extraction	NARMAX, Model Structure Selection	Extracting useful information of each term from sampled model structures.
[76]	Feature selection and Classification	Polynomial expansion	Classification approach adopted from the nonlinear model identification framework
[38]	Feature selection	Neighbourhood Component Analysis (NCA)	A novel nearest neighbor-based feature weighting algorithm

[39]	Feature selection and Classification	NCA, KNN classification	A novel method for learning a Mahalanobis Distance measure to be used in the KNN classification algorithm
[40]	Feature selection and Classification	NCA, Hidden Markov model	Bearing fault diagnosis and fault severity classification
[41]	Classification	Machine Learning	Review Machine learning methods for wind turbine condition monitoring

533

REFERENCES

1. Pérez, J.M.P.; Márquez, F.P.G.; Hernández, D.R. Economic viability analysis for icing blades detection in wind turbines. *Journal of Cleaner Production* **2016**, *135*, 1150-1160.
2. Márquez, F.P.G.; Pérez, J.M.P.; Marugán, A.P.; Papaelias, M. Identification of critical components of wind turbines using fta over the time. *Renewable Energy* **2016**, *87*, 869-883.
3. Márquez, F.P.G.; Pedregal, D.J.; Roberts, C. New methods for the condition monitoring of level crossings. *International Journal of Systems Science* **2015**, *46*, 878-884.
4. Pliego Marugán, A.; García Márquez, F.P.; Pinar Pérez, J.M. Optimal maintenance management of offshore wind farms. *Energies* **2016**, *9*, 46.
5. Abdelgawad, A.; Yelamarthi, K. Internet of things (iot) platform for structure health monitoring. *Wireless Communications and Mobile Computing* **2017**, *2017*.
6. Zhou, Y.; He, F.; Qiu, Y. Optimization of parallel iterated local search algorithms on graphics processing unit. *The Journal of Supercomputing* **2016**, *72*, 2394-2416.
7. Gomez Munoz, C.; De la Hermosa Gonzalez-Carrato, R.; Trapero Arenas, J.; Garcia Marquez, F. A novel approach to fault detection and diagnosis on wind turbines. *Global Nest Journal* **2014**, *16*, 1029-1037.
8. Ramirez, I.S.; Muñoz, C.Q.G.; Marquez, F.P.G. In *A condition monitoring system for blades of wind turbine maintenance management*, Proceedings of the tenth international conference on management science and engineering management, 2017; Springer: pp 3-11.
9. McGugan, M.; Pereira, G.; Sørensen, B.F.; Toftegaard, H.; Branner, K. Damage tolerance and structural monitoring for wind turbine blades. *Phil. Trans. R. Soc. A* **2015**, *373*, 20140077.
10. McGugan, M. Design of wind turbine blades. In *Mare-wint*, Springer: 2016; pp 13-24.
11. Pérez-Ortiz, M.; Jiménez-Fernández, S.; Gutiérrez, P.A.; Alexandre, E.; Hervás-Martínez, C.; Salcedo-Sanz, S. A review of classification problems and algorithms in renewable energy applications. *Energies* **2016**, *9*, 607.
12. Joshuva, A.; Sugumaran, V. Fault diagnostic methods for wind turbine: A review. *ARPN Journal of Engineering and Applied Sciences* **2016**, *11*, 4654_4668.
13. Dervilis, N.; Choi, M.; Taylor, S.; Barthorpe, R.; Park, G.; Farrar, C.; Worden, K. On damage diagnosis for a wind turbine blade using pattern recognition. *Journal of sound and vibration* **2014**, *333*, 1833-1850.
14. Pratumnopharat, P.; Leung, P.S.; Court, R.S. Wavelet transform-based stress-time history editing of horizontal axis wind turbine blades. *Renewable Energy* **2014**, *63*, 558-575.
15. Altabey, W.A.; Noori, M. An extensive overview of lamb wave technique for detecting fatigue damage in composite structures. *Industrial and Systems Engineering* **2017**, *2*, 20.
16. Feklistova, L.; Hein, H. Delamination identification using machine learning methods and haar wavelets. *Computer Assisted Methods in Engineering and Science* **2017**, *19*, 351-360.
17. Pavlopoulou, S.; Worden, K.; Soutis, C. Novelty detection and dimension reduction via guided ultrasonic waves: Damage monitoring of scarf repairs in composite laminates. *Journal of Intelligent Material Systems and Structures* **2016**, *27*, 549-566.
18. Jiménez, A.A.; Muñoz, C.Q.G.; Márquez, F.P.G. Machine learning for wind turbine blades maintenance management. *Energies* **2017**, *11*, 1-16.
19. Papaelias, M.; Cheng, L.; Kogia, M.; Mohimi, A.; Kappatos, V.; Selcuk, C.; Constantinou, L.; Muñoz, C.Q.G.; Marquez, F.P.G.; Gan, T.-H. Inspection and structural health monitoring techniques for concentrated solar power plants. *Renewable Energy* **2016**, *85*, 1178-1191.
20. Muñoz, C.Q.G.; Marquez, F.P.G.; Liang, C.; Maria, K.; Abbas, M.; Mayorkinos, P. In *A new condition monitoring approach for maintenance management in concentrate solar plants*, Proceedings of the ninth international conference on management science and engineering management, 2015; Springer: pp 999-1008.
21. Jiménez, A.A.; Muñoz, C.Q.G.; Marquez, F.P.G.; Zhang, L. In *Artificial intelligence for concentrated solar plant maintenance management*, Proceedings of the Tenth International Conference on Management Science and Engineering Management, 2017; Springer: pp 125-134.
22. Muñoz, C.Q.G.; Marquez, F.P.G.; Liang, A.A.J.; Cheng, M.K.; Mohimi, A.; Papaelias, M.; Mohimi, A.; Cheng, L. Fault detection and diagnosis employing the electromagnetic sensors emat.
23. Raghavan, A.; Cesnik, C.E. Review of guided-wave structural health monitoring. *Shock and Vibration Digest* **2007**, *39*, 91-116.
24. Yu, L.; Bao, J.; Giurgiutiu, V. In *Signal processing techniques for damage detection with piezoelectric wafer active sensors and embedded ultrasonic structural radar*, Smart Structures and Materials, 2004; International Society for Optics and Photonics: pp 492-503.

25. Rizzo, P.; di Scalea, F.L. In *Discrete wavelet transform to improve guided-wave-based health monitoring of tendons and cables*, Smart Structures and Materials, 2004; International Society for Optics and Photonics: pp 523-532.
26. Hamming, R.W. Digital filters. 3rd. Prentice Hall: 1989.
27. García Márquez, F.P.; García‐Pardo, I.P. Principal component analysis applied to filtered signals for maintenance management. *Quality and Reliability Engineering International* **2010**, *26*, 523-527.
28. Márquez, F.P.G.; Muñoz, J.M.C. A pattern recognition and data analysis method for maintenance management. *International Journal of Systems Science* **2012**, *43*, 1014-1028.
29. Zhang, T.; Chen, W.; Li, M. Ar based quadratic feature extraction in the vmd domain for the automated seizure detection of eeg using random forest classifier. *Biomedical Signal Processing and Control* **2017**, *31*, 550-559.
30. Leontaritis, I.; Billings, S.A. Input-output parametric models for non-linear systems part i: Deterministic non-linear systems. *International journal of control* **1985**, *41*, 303-328.
31. Worden, K.; Surace, C.; Becker, W. Uncertainty bounds on higher-order frfs from gaussian process narx models. *Procedia Engineering* **2017**, *199*, 1994-2000.
32. Rai, A.; Upadhyay, S. The use of md-cumsum and narx neural network for anticipating the remaining useful life of bearings. *Measurement* **2017**, *111*, 397-410.
33. Hussain, S.; Gabbar, H.A. Vibration analysis and time series prediction for wind turbine gearbox prognostics. *Int. J. Progn. Health Manage* **2013**, *4*, 69-79.
34. Leite, G.d.N.P.; Araújo, A.M.; Rosas, P.A.C. Prognostic techniques applied to maintenance of wind turbines: A concise and specific review. *Renewable and Sustainable Energy Reviews* **2017**.
35. Liu, W.; Li, C.; Liu, Y.; Wu, Q. Predictive control of wind turbine for load reduction during ramping events. *International Journal of Electrical Power & Energy Systems* **2017**, *93*, 135-145.
36. Men, Z.; Yee, E.; Lien, F.-S.; Wen, D.; Chen, Y. Short-term wind speed and power forecasting using an ensemble of mixture density neural networks. *Renewable Energy* **2016**, *87*, 203-211.
37. Jiménez, A.A.; Márquez, F.P.G.; Moraleda, V.B.; Muñoz, C.Q.G. Linear and nonlinear features and machine learning for wind turbine blade ice detection and diagnosis. *Renewable Energy* **2019**, *132*, 1034-1048.
38. Yang, W.; Wang, K.; Zuo, W. Neighborhood component feature selection for high-dimensional data. *JCP* **2012**, *7*, 161-168.
39. Goldberger, J.; Hinton, G.E.; Roweis, S.T.; Salakhutdinov, R.R. In *Neighbourhood components analysis*, Advances in neural information processing systems, 2005; pp 513-520.
40. Zhou, H.; Chen, J.; Dong, G.; Wang, H.; Yuan, H. Bearing fault recognition method based on neighbourhood component analysis and coupled hidden markov model. *Mechanical Systems and Signal Processing* **2016**, *66*, 568-581.
41. Stetco, A.; Dinmohammadi, F.; Zhao, X.; Robu, V.; Flynn, D.; Barnes, M.; Keane, J.; Nenadic, G. Machine learning methods for wind turbine condition monitoring: A review. *Renewable Energy* **2018**.
42. Gómez Muñoz, C.Q.; García Márquez, F.P. A new fault location approach for acoustic emission techniques in wind turbines. *Energies* **2016**, *9*, 40.
43. de la Hermosa González, R.R.; Márquez, F.P.G.; Dimlaye, V.; Ruiz-Hernández, D. Pattern recognition by wavelet transforms using macro fibre composites transducers. *Mechanical Systems and Signal Processing* **2014**, *48*, 339-350.
44. de la Hermosa González, R.R.; Márquez, F.P.G.; Dimlaye, V. Maintenance management of wind turbines structures via mfcs and wavelet transforms. *Renewable and Sustainable Energy Reviews* **2015**, *48*, 472-482.
45. Eristi, H. Fault diagnosis system for series compensated transmission line based on wavelet transform and adaptive neuro-fuzzy inference system. *Measurement* **2013**, *46*, 393-401.
46. Daubechies, I. *Ten lectures on wavelets*. SIAM: 1992.
47. Muñoz, C.Q.G.; Jiménez, A.A.; Márquez, F.P.G. Wavelet transforms and pattern recognition on ultrasonic guides waves for frozen surface state diagnosis. *Renewable Energy* **2018**, *116*, 42-54.
48. Chiementin, X.; Kilundu, B.; Rasolofondraibe, L.; Crequy, S.; Pottier, B. Performance of wavelet denoising in vibration analysis: Highlighting. *Journal of Vibration and Control* **2012**, *18*, 850-858.
49. Stoica, P.; Moses, R.L. *Introduction to spectral analysis*. Prentice hall Upper Saddle River: 1997; Vol. 1.
50. Ahn, S.K.; Reinsel, G.C. Estimation of partially nonstationary vector autoregressive models with seasonal behavior. *Journal of Econometrics* **1994**, *62*, 317-350.

51. Zhang, L.; Li, K.; Bai, E.-W.; Irwin, G.W. Two-stage orthogonal least squares methods for neural network construction. *IEEE transactions on neural networks and learning systems* **2015**, *26*, 1608-1621.
52. Chen, S.; Billings, S. Representations of non-linear systems: The narmax model. *International Journal of Control* **1989**, *49*, 1013-1032.
53. Loh, W.-Y.; Shih, Y.-S. Split selection methods for classification trees. *Statistica sinica* **1997**, 815-840.
54. Deconinck, E.; Zhang, M.H.; Coomans, D.; Vander Heyden, Y. Classification tree models for the prediction of blood− brain barrier passage of drugs. *Journal of chemical information and modeling* **2006**, *46*, 1410-1419.
55. Nardi, D.; Lampani, L.; Pasquali, M.; Gaudenzi, P. Detection of low-velocity impact-induced delaminations in composite laminates using auto-regressive models. *Composite Structures* **2016**.
56. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. In *A training algorithm for optimal margin classifiers*, Proceedings of the fifth annual workshop on Computational learning theory, 1992; ACM: pp 144-152.
57. Milgram, J.; Cheriet, M.; Sabourin, R. In *"One against one" or "one against all": Which one is better for handwriting recognition with svms?*, Tenth International Workshop on Frontiers in Handwriting Recognition, 2006; Suvisoft.
58. Friedman, J.; Hastie, T.; Tibshirani, R. *The elements of statistical learning*. Springer series in statistics New York: 2001; Vol. 1.
59. Dietterich, T.G. Ensemble methods in machine learning. *Multiple classifier systems* **2000**, *1857*, 1-15.
60. Breiman, L. Random forests. *Machine learning* **2001**, *45*, 5-32.
61. Lobo, J.M.; Jiménez‐Valverde, A.; Real, R. Auc: A misleading measure of the performance of predictive distribution models. *Global ecology and Biogeography* **2008**, *17*, 145-151.
62. Bradley, A.P. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern recognition* **1997**, *30*, 1145-1159.
63. Demšar, J. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine learning research* **2006**, *7*, 1-30.
64. Garcia, S.; Herrera, F. An extension on``statistical comparisons of classifiers over multiple data sets"for all pairwise comparisons. *Journal of Machine Learning Research* **2008**, *9*, 2677-2694.
65. Conover, W.J.; Iman, R.L. Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician* **1981**, *35*, 124-129.
66. Blair, C.R.; Hochberg, Y. Improved bonferroni procedures for testing overall and pairwise homogeneity hypotheses. *Journal of Statistical Computation and Simulation* **1995**, *51*, 281-289.
67. Holm, S. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics* **1979**, 65-70.
68. Rouder, J.N.; Speckman, P.L.; Sun, D.; Morey, R.D.; Iverson, G. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic bulletin & review* **2009**, *16*, 225-237.
69. Chopra, I. Review of state of art of smart structures and integrated systems. *AIAA journal* **2002**, *40*, 2145-2187.
70. Lopez, I.; Sarigul-Klijn, N. A review of uncertainty in flight vehicle structural damage monitoring, diagnosis and control: Challenges and opportunities. *Progress in Aerospace Sciences* **2010**, *46*, 247-273.
71. Jardine, A.K.; Lin, D.; Banjevic, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing* **2006**, *20*, 1483-1510.
72. Hunter, N.F.; Worden, K. Using statistical pattern recognition techniques. *Engineering Analysis* **2001**, *1000*, C926.
73. Hong, X.; Mitchell, R.J.; Chen, S.; Harris, C.J.; Li, K.; Irwin, G.W. Model selection approaches for non-linear system identification: A review. *International journal of systems science* **2008**, *39*, 925-946.
74. KORENBERG, M.; Billings, S.; Liu, Y.; McIlroy, P. Orthogonal parameter estimation algorithm for non-linear stochastic systems. *International Journal of Control* **1988**, *48*, 193-210.
75. Falsone, A.; Piroddi, L.; Prandini, M. A randomized algorithm for nonlinear model structure selection. *Automatica* **2015**, *60*, 227-238.
76. Brankovic, A.; Falsone, A.; Prandini, M.; Piroddi, L. Randomised algorithm for feature selection and classification. *arXiv preprint arXiv:1607.08400* **2016**.